Junseok Park

A computational scientist, data scientist, or bioinformatician with expertise in computer science, skilled in developing machine learning models, designing large-scale data processing pipelines, and performing statistical analyses of next-generation sequencing and multi-omic experiments from patient cohorts—ranging from cellular data to DNA and RNA sequence samples. Highly collaborative, I have contributed to projects ranging from cancer immunology and amyotrophic lateral sclerosis to dementia and neurodegeneration in ataxia-telangiectasia.

Email	junseokpark.kr@gmail.com
ORCID	0000-0003-2312-9440
Website	junseokpark.github.io
Linkedin	junseok-park-ph-d-42a171a8

Education

2016–2020 **PhD** program in **Bioinformatics** – KAIST (Korea Advanced Institute of Science and Technology)

Thesis: Reliable data collection in participatory trials to assess digital healthcare applications
Supervisor: Dr. Doheon Lee

2014–2016 **MSc** in **Bioinformatics** – KAIST (Korea Advanced Institute of Science and Technology)

Thesis: A system development for efficient clinical trials in clinical research: CORUS
Supervisor: Dr. Doheon Lee and Dr. Kwang hyung Lee

2000–2007 **BSc** in Electic and Computer Engineering – Chonnam National University

Specialist: **Computer Science**
Minor: **Electronic and Electrical Engineering**
Advisor: Dr. Yonggwan Won

Publications

- **Highlighted works**

Boram Lee*, **Junseok Park***, Adam Voshall, Yangmin Gan, Eduardo Maury, Yeeok Kang et al. *Pan-cancer analysis reveals multifaceted roles of retrotransposon-fusion RNAs.* Nature Communication (2024): In revision.
10.1101/2023.10.16.562422, github.com/junseokpark/rtea

Junseok Park, Eduardo Maury, Changhoon Oh, Donghoon Shin, Danielle Denisko, Eunjung Alice Lee. *Genomic data processing with GenomeFlow.* Accepted to BMC Bioinformatics (2024). github.com/junseokpark/genomeflow

Miaomiao Tan*, Zhinan Lin*, Zhuofu Chen, **Junseok Park**, Ziting He, Haonan Zhou et al. *Image-based DNA Sequencing Encoding for Detecting Low-Mosaicism Somatic Mobile Element Insertions.* Nature Method (2024): In revision. 10.1101/2024.11.07.619809

SMaHT Data Analysis Center (DAC), *A synthetic mosaicism benchmark from multiple-individuals in a mixed sample.* In-preparation. github.com/junseokpark/SMaHT

Jeffrey J. Widrick, Matthias R. Lambert, Felipe de Souza Leite, Youngsook Lucy Jung, **Junseok Park**, James R. Conner et al. *Kinematic phenotyping of dystrophic zebrafish larvae*, Science advances (2024): In revision. 10.1101/2024.12.05.627004

Lai, Jenny, Didem Demirbas, Junho Kim, Ailsa M. Jeffries, Allie Tolles, **Junseok Park** et al. *ATM-deficiency-induced microglial activation promotes neurodegeneration in ataxia-telangiectasia.* Cell reports (2024): 43, no. 1. 10.1016/j.celrep.2023.113622

Zinan Zhou, Junho Kim, August Yue Huang, Matthew Nolan, **Junseok Park**, Ryan Doan et al. *Somatic Mosaicism in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia Reveals Widespread Degeneration from Focal Mutations.* [Preprint] bioRxiv (2023). 10.1101/2023.11.30.569436

*Co-first authors

- **Journal and Preprints**

  Sunjae Lee, Victoria Meslier, Gholamreza Bidkhori, Fernando Garcia-Guevara, Lucie Etienne-Mesmin, Frederick Clasen, **Junseok Park** et al. *Transient colonizing microbes promote gut dysbiosis and functional impairment.* npj Biofilms and Microbiomes (2024): 10,80. 10.1038/s41522-024-00561-1

  Saeed Shoaie, Sunjae Lee, Mathieu Almeida, Gholamreza Bidkhori, Nicolas Pons, Florian Onate, ..., **Junseok Park** et al. *Archive Global and temporal state of the human gut microbiome in health and disease.* Research Square (2021). 10.21203/rs.3.rs-339282/v1

  Junseok Park, Kwangmin Kim, Seongkuk Park, Woochang Hwang, Sunyong Yoo, Gwansu Yi, Doheon Lee. *An interactive retrieval system for clinical trial studies with context-dependent protocol elements.* PloS one (2020): 15.9:e0238290. 10.1371/journal.pone.0238290, github.com/junseokpark/clips

  Stefania Vaga, Sunjae Lee, Boyang Ji, Anna Andreasson, Nicholas J Talley, Lars Agréus, ..., **Junseok Park** et al. *Compositional and functional differences of the mucosal microbiota along the intestine of healthy individuals.* Scientific Reports (2020): 10, 14977 10.1038/s41598-020-71939-2

  Junseok Park, Seongkuk Park, Kwangmin Kim, Gwangmin Kim, Jaegyun Jung, Sunyong Yoo et al. *Reliable Data Collection in Participatory Trials to Assess Digital Healthcare Applications.* IEEE Access (2020): 79472-79490. 10.1109/ACCESS.2020.2985122, github.com/junseokpark/corus

  Junseok Park, Kwangmin Kim, Woochang Hwang, and Doheon Lee. *Concept embedding to measure semantic relatedness for biomedical information ontologies.* Journal of biomedical informatics (2019): 94:103182. 10.1016/j.jbi.2019.103182

  Yoo, Sunyong, Kyungrin Noh, Moonshik Shin, **Junseok Park**, Kwang-Hyung Lee, Hojung Nam, and Doheon Lee. *In silico profiling of systemic effects of drugs to predict unexpected interactions.* Scientific reports (2018): 8, no. 1: 1612. 10.1038/s41598-018-19614-5

  Yu, Hasun, Sungji Choo, **Junseok Park**, Jinmyung Jung, Yeeok Kang, and Doheon Lee. *Prediction of drugs having opposite effects on disease genes in a directed network.* BMC systems biology (2016): vol. 10, no. 1, p. S2. 10.1186/s12918-015-0243-2

  Kim, Docyong, Jaehyun Lee, Sunjae Lee, **Junseok Park**, and Doheon Lee. *Predicting unintended effects of drugs based on off-target tissue effects.* Biochemical and biophysical research communications (2016): 469, no.3: 399-404 10.1016/j.bbrc.2015.11.095

- **Conference proceedings**

  Junseok Park, Seongkuk Park, Kwangmin Kim, and Doheon Lee. *CORUS: Blockchain-Based Trustworthy Evaluation System for Efficacy of Healthcare Remedies.* IEEE International Conference on Cloud Computing Technology and Science (CloudCom) (2018): 181-184. 10.1109/CloudCom2018.2018.00044

Work Experience

Aug. 2020 – present **Lee Lab** – Boston Children's Hospital and Harvard Medical School, Boston, MA, USA
Research Fellow (PI: Dr. Alice Eunjung Lee)

I am designing a deep learning model to identify correlations between transposon elements (TEs) and alternative splicing regions, aiming to uncover potential connections that could enhance our understanding of gene regulation. Additionally, I developed a tool with the potential to significantly accelerate the detection of TE-related novel oncogenes and biomarkers for immunotherapy, which could have impactful applications in cancer research. To support these efforts, I also created a parallel processing architecture capable of handling large-scale genomic samples using cloud computing, enabling efficient analysis and scalability for future research endeavors.

Jan. 2012 – Feb 2014 **Korea Research Institute of Chemical Technology (KRICT)** – Daejeon, Republic of Korea
Technical Manager

I developed a chemical reagent management and analysis system to streamline the organization and tracking of reagents, enhancing overall laboratory efficiency. Additionally, I planned long-term strategies for KRICT, where I played a key role in developing the IT infrastructure to support research achievements and recording systems, ensuring a more structured and efficient data management process.

Nov 2011 – Dec. 2011 **Seoul National University Hospital** – Seoul, Republic of Korea
Planning Manager

I led the development of a next-generation hospital management system, which included integrating Electronic Health Records (EHR), Electronic Medical Records (EMR), and Personal Health Records (PHR) to enhance healthcare data management and improve patient care.

Apr. 2011 – Nov. 2011 **Korea BIO-IT Foundry Gwang-ju Centre** – Gwangju, Republic of Korea
Researcher (PI: Dr. Yonggwan Won)

I developed a high-speed additive reagent injection and inspection system to improve precision and efficiency in chemical processes. Additionally, I helped create a smart home experience space integrated with various bio-IT prototype products, providing a connected environment to showcase and test innovative home technologies.

Jan. 2007 – Mar. 2010 **SKTelecom** – Gwangju/Seoul, Republic of Korea
Manager

I enhanced the system efficiency of WCDMA networks, optimizing their performance for better data transmission and network reliability. Additionally, I improved the stability of WCDMA systems by implementing updates such as the ISP (Inter-Section Paging) function, ensuring more consistent and reliable connectivity for users.

Mar. 2001 – Apr. 2003 **Republic of Korea Army** – 31 Division, Republic of Korea
Sergeant, Honorable discharge

I developed ocean protection radar systems and networks along the Korean coast to enhance coastal surveillance and maritime safety. Additionally, I contributed to the development of communication networks for thermal observation devices (TODs), enabling more effective monitoring and data transmission for critical security and environmental applications.

Presentations

- **Invited**

  "Reliable data collection in participatory trials to assess digital healthcare applications", NEBS-Seoul National University Hospital Meeting, Whitehead Institute, MIT, USA. Jun 15, 2024

  "Pan-Cancer Analysis: Unveiling the Functions of Retrotransposon Fusion RNAs", NEBS Annual Conference 2024, Tosteson Medical Education Center (TMEC) Harvard Medical School, Boston, MA, USA. May 18, 2024

  " Genomic data processing with GenomeFlow", DTMBIO 2022, Hilton Walkoloa Village, Hawaii, USA. Dec 20, 2022.

  "GenomeFlow-Performance Test Preparation", Technical Meet-Up, Google, Cambridge, MA, USA. Mar 1, 2022.

  "Bioinformatics based on computer science", Invited Talks for Systems Biology Lab (Dr. Saeed Shoaie), Guy's Campus, King's College London, London, United Kingdom. Sep 2, 2019

  "Blockchain for Science and Research", Invited Talks for Park Lab (Dr. Peter Park), Countway Library, Harvard Medical School, Boston, MA, USA. Jun 13, 2019

  "Blockchain for Science and Research" Amazon Web Service Public Sector Summit, Walter E. Washington Convention Centre, Washington DC, USA. June 11, 2019. https://youtu.be/BjZObg_A3Jg

  "Collaborative Research for us" Amazon Web Service, Seoul Summit, Coex, Seoul, Republic of Korea. Apr 17, 2019. https://youtu.be/Y7CRxClW5Tc

  "CORUS: a blockchain-based trustworthy evaluation system for efficacy of healthcare remedies", Cloudcom 2018, Hilton Cyprus, Nicosia, Cyprus. Dec 12, 2018.

- **Conference, poster presentations**

  "Pan-cancer analysis reveals roles of retrotransposon-fusion RNAs" ASHG 2023, Washington DC, Walter E. Washington Convention Centre, USA. Nov 5, 2023

  "Functional Investigation of Somatic Variants in Pediatric Epilepsy Using Single-Cell DNA/RNA Assays from Patient Derived Single Cells (P3-9.009)". American Academy of Neurology 2023, Boston, MA, USA. Apr 23, 2023

  "An artificial intelligence based, markerless motion capture approach for quantifying motor deficits of dystrophic zebrafish larvae". Clinical Scientific Conference (2023): S30, Nov 5, 2023

  "Parallel Text Mining for Extracting Co-Occurrence Terms from Massive Literature". International Conference on Convergence Content 2018, International Convention Center Jeju, Republic of Korea. Dec 18, 2018

  "Advanced UMLS semantic relatedness measures on concept vectors". IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2017, Kansas City, MO, USA. Nov 13-16, 2017

  "A bioinformatics system for searching Co-Occurrence based on Co-Operational Formation with Advanced Method (COCOFAM)". International Society for Computational Biology(ISMB) 2015, Convention Centre, Dublin, Ireland. July 13, 2015

- **Conference, attendance**

  The Festival of Genomics and Biodata, Boston Convention and Exhibition Center (BCEC), Boston, MA, USA. June 12-13, 2024

  Neural Information Processing Systems Online Conference, Dec 6-12, 2020

  34th IEEE International Conference on Data Engineering, Apr 16-19, Conservatoire National des Arts et Métiers (CNAM), Paris, France. Apr 16-19, 2018

  Strata + Hadoop World 2016: O'Reilly Big Data Conference, San Jose Convention Center, San Jose, CA, USA. Mar 29-31, 2016

- **Workshop**

  EMBO(European Molecular Biology Organization) Workshop, Autophagy: From molecular principles to human diseases, Crieff, Scotland, United Kingdom, Aug 26-30, 2019

- **Forum**

  Hyperledger Global Forum 2018, Basel Congress Center, Basel, Switzerland. Dec 12-15, 2018

- **Symposium**

  6th Uppsala Transposon Symposium, Friessalen, Evolutionary Biology Centre, Uppsala, Sweden. Oct 26-28, 2022

  Milner Symposium 2017, West Road Concert Hall, Cambridge, United Kingdom. Oct 2, 2017

  Healthcare Information and Management Systems Society (HIMSS) 2017, Orlando, FL, USA. Feb 19-23, 2017

  Bio Synergy 2016 Symposium,TNO Zeist/Leiden, Netherland. Aug 23-27, 2016

Teaching

Spring 2019 Teaching Fellow for **BiS301 Bioengineering Laboratory I - Lab 10**, KAIST
Course covers C++ programming with gene database.
Topics Include: Decision Tree, SQL Database, Probability, Regression, and Classification.
Professors: Dr. Yoonkey Nam and Dr. Doheon Lee

Fall 2017 Teaching Fellow for **BiS732 Bio-Network**, KAIST
Course covers systems biology.

Topics Include: Metabolic Pathways, Signal Transduction Pathways, Regulation Networks, Graph Theory, Boolean Networks, Bayesian Networks, Bio-Network Modeling, and Formal Representation Tools.
Professor: Dr. Doheon Lee

Spring 2018, Spring 2017 Teaching Fellow for **BiS332 Bio Data Mining**, KAIST
Course covers fundamental and essential understanding of how modern electronic computers work.
Topics Include: Bioinformatics algorithms, Machine Learning, Data Mining, Biomedical Information Systems, Personalized Medicine, System Design, and Implementation.
Professor: Dr. Doheon Lee

Fall 2016 Teaching Fellow for **BiS232 Bio-Data Structure**, KAIST
Course explores practical implementation of ideas in biology and brain science through C++ programming and fundamental data structures.
Topics Include: Cloud computing architecture, Decision Tree, SQL Database, probability, regression, and classification.
Professor: Dr. Sangwan Lee

Awards and Honours

2024 NEBS-Tomocube Award Presentation (NEBS-Tomocube)
2021–2024 Basic Science Research Funding (National Research Fund, Republic of Korea)
2022 AnVIL Cloud Credits Continued Program (National Human Genome Research Institute, USA)
2016–2020 Korean Government Scholarship, Ph.D. (KAIST)
2019 KIA Scholarship (KAIST)
2017 ICC Scholarship (KAIST)
2009 SKMS Practice prize - Regional office (SKTelecom)
2009 SKMS Practice prize - Division (SKTelecom)
2005–2006 Full Scholarship (Chonnam National University)

Organizations

2021–Present New England Bioscience Society
2017-2020 IEEE Young Professionals

Patents

Jan 2022 Method and Apparatus for Performance Evaluating of Healthcare Applications
KR20210081545A, https://patents.google.com/patent/KR20210081545A
Jan 2020 Method and Apparatus for Data Managing for Clinical Trial
KR20190094729A, https://patents.google.com/patent/KR20190094729A
Jul 2017 A Method for Searching Co-Occurrence Based on Co-Operational Formation
KR20160149619A, https://patents.google.com/patent/KR20160149619A

Certificates

2011–Present **Network Management** – Information & Communication Qualification Association of Korea
2011–Present **Engineer Information Processing** – Human Resources Development Service of Korea
2009–2010 **Six Sigma Green Belt** – SKTelecom, Seoul, Korea
2007–2009 **Cisco Certified Network Associate** – CISCO, Seoul, Korea

Other

Skills: Deep learning, Machine learning, bioinformatics, Cloud computing, Genomics, Python, R, Scala, C++, Java
Operating systems: Linux (CentOS), MacOS, Windows
Languages: Korean (fluent), English (fluent), Japanese (fluent)

References

Alice Eunjung Lee, Ph.D Harvard Medical School – Associate Professor
https://compgen.hms.harvard.edu/people/eunjung-alice-lee
Contact : ealee@childrens.harvard.edu

Doheon Lee, Ph.D KAIST – Endowed Chair Professor
https://sites.google.com/view/doheonlee
Contact : dhlee@kaist.ac.kr